

OCEANTEA: EXPLORING OCEAN-DERIVED CLIMATE DATA USING MICROSERVICES

Arne N. Johanson¹, Sascha Flögel², Wolf-Christian Dullo², Wilhelm Hasselbring¹

Abstract—Ocean observation systems gather an increasing amount of climate-relevant time series data. To interactively explore and analyze such high-dimensional datasets, we developed the software OceanTEA. Our open-source tool leverages modern web technology to support interactive data visualization, spatial analysis of current patterns, and temporal pattern discovery via machine learning methods. The microservice architecture of OceanTEA ensures a maintainable implementation that seamlessly scales from desktop computers to cloud computing infrastructure.

I. MOTIVATION

Ocean observation systems, such as the global array of more than 3000 free-drifting Argo floats belonging to the Global Ocean Observing System [1] or the modular ocean laboratory MoLab [2, 3], produce an increasing amount of time series data. Both statistical data mining techniques and manual exploration via visualization are necessary for oceanographers and climatologists to extract scientific knowledge from such vast datasets. For this purpose, we developed the software OceanTEA, which leverages modern web technology to support scientists in interactively exploring and analyzing high-dimensional datasets. By relying on a microservice architecture [4, 5], OceanTEA can not only be deployed on desktop computers but also on cloud computing infrastructure with built-in scalability. Making data available on the web can be useful for scientists collaborating on exploring a dataset (e.g., with limited access within an institute) as well as for providing interactive visualizations along with journal or conference publications. Since it has been shown that papers which feature published data receive higher citation counts [6], an interactive visualization of such data with OceanTEA could further improve the impact of a publication.

Corresponding author: A. Johanson, arj@informatik.uni-kiel.de ¹Software Engineering Group, Kiel University, Germany
²GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

The OceanTEA source code (along with a live demo of the tool) is available on GitHub¹ under the Apache 2.0 license [7].

A tool related to OceanTEA is Ocean Data View (ODV) [8], which is a proprietary (i.e., closed-source) desktop-only application used to produce a wide range of *static* figures from oceanographic datasets.

II. OCEANTEA

A screenshot of the web interface of OceanTEA (short for *Oceanographic Time Series Exploration and Analysis*) is shown in Figure 1. The user interface is divided into four views (times series management, data exploration, spatial analysis, and temporal pattern discovery), which can be accessed via the tabs at the top of the page. The data exploration view (Figure 1) features options to filter the time series to be displayed according to:

- 1) study region
- 2) measurement device
- 3) measurement parameter (e.g., temperature)
- 4) depth range (multiple ranges are possible)

Furthermore, measurement stations can directly be selected via an interactive map displaying satellite images of the Earth's surface (provided by Google Maps [9]).

OceanTEA supports both univariate time series (e.g., temperature measured at a single site) and multivariate series (e.g., current direction and magnitude in several depth bins in the water column measured by an acoustic Doppler current profiler (ADCP)). Multivariate time series of currents (direction and magnitude) can be sliced along adjustable depth levels (see the right plot in Figure 1).

The interactive plots of OceanTEA are implemented using the CavaPlot [10] library built on top of D3.js [11]. The user can zoom into the plots and pan the axes (also by using touch gestures on devices that support them). At a high zoom level, the individual data points are displayed and tooltips are shown when

¹<https://github.com/a-johanson/oceantea>

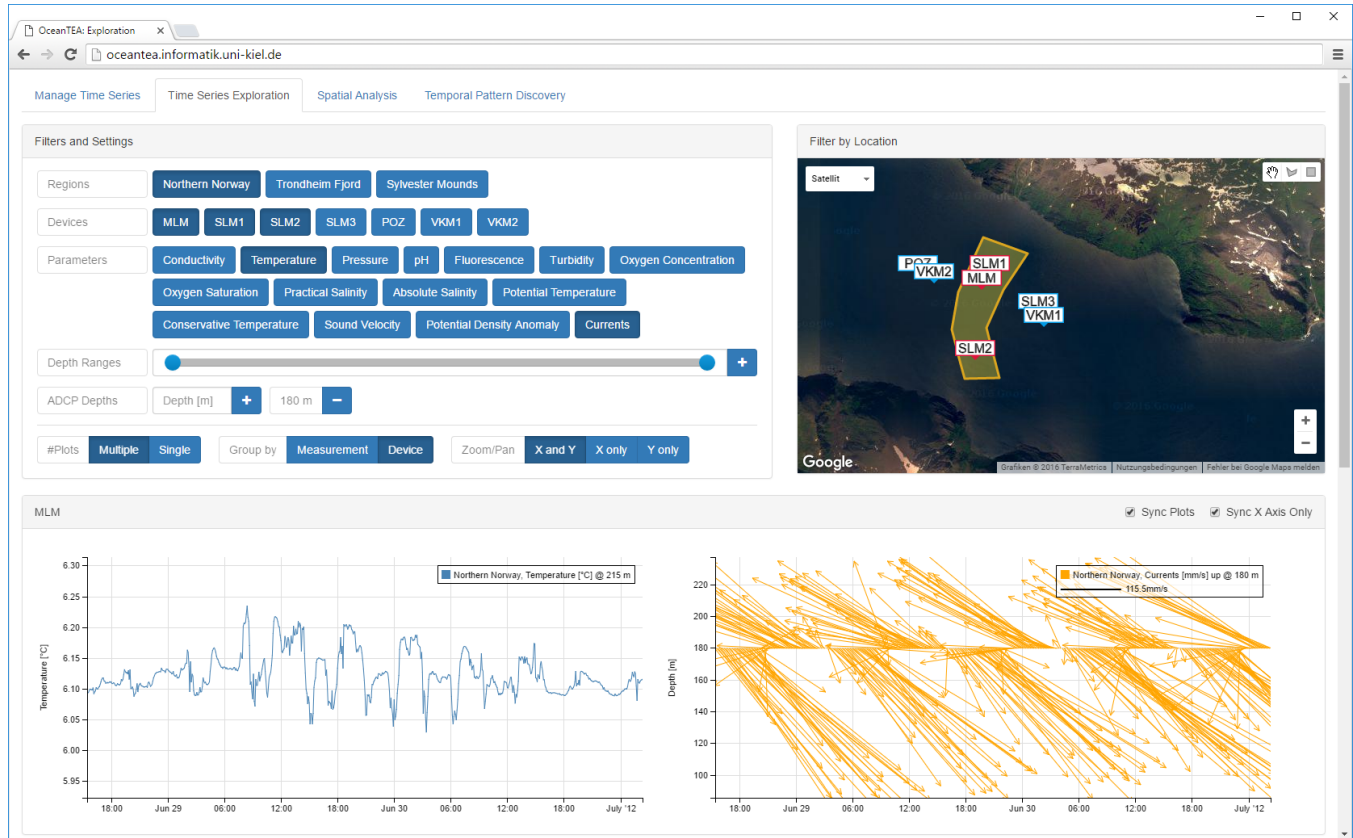


Figure 1. The data exploration view of OceanTEA.

the user hovers over the points (or touches them). The plots can be arranged by measurement parameter or by measurement device. It is possible to synchronize the axes of multiple plots and to join multiple graphs in a single plot.

In the management tab, time series can be added (e.g., from comma-separated values (CSV) files) and deleted. OceanTEA provides automatic unit conversion according to TEOS-10 [12] for several important oceanographic parameters (such as from *in situ* temperature to conservative temperature).

III. MICROSERVICE ARCHITECTURE

The microservice architecture pattern [4, 5] partitions a software system into a set of so-called microservices. A microservice is a small, self-contained application that can be deployed independently and has a single responsibility [13]. In this context, small means that its complexity is low enough to be understood by a small team or even a single developer. That microservices are self-contained implies that they do not share code or database schemas with each other. In particular, each microservice can be implemented using the programming languages, middleware, and data

storage facilities that suit the task of the service best (polyglot programming and persistence). As the whole software system is divided into microservices according to domain functionality (in the sense of bounded contexts in domain-driven design [14]), each service only has a single functional responsibility. Transaction-less communication—e.g., via RESTful protocols such as HTTP—is employed to coordinate tasks between the individual services.

While microservices incur the drawback of having to handle the additional complexity of distributed systems (e.g., ensuring fault tolerance), they provide the advantage of good maintainability and scalability [15]. As the complexity of a microservice is low, maintaining its code is easier than that of a large monolithic application (making it a feasible option to re-implement the whole service if necessary). Since microservices are self-contained and can be deployed independently, they can also be scaled independently as it is required by the current workload on the software system [15, 16, 17].

Figure 2 shows the microservice architecture of OceanTEA. The OceanTEA client, which contains the user interface and runs in the user's web browser, communicates with the server-side part of OceanTEA via

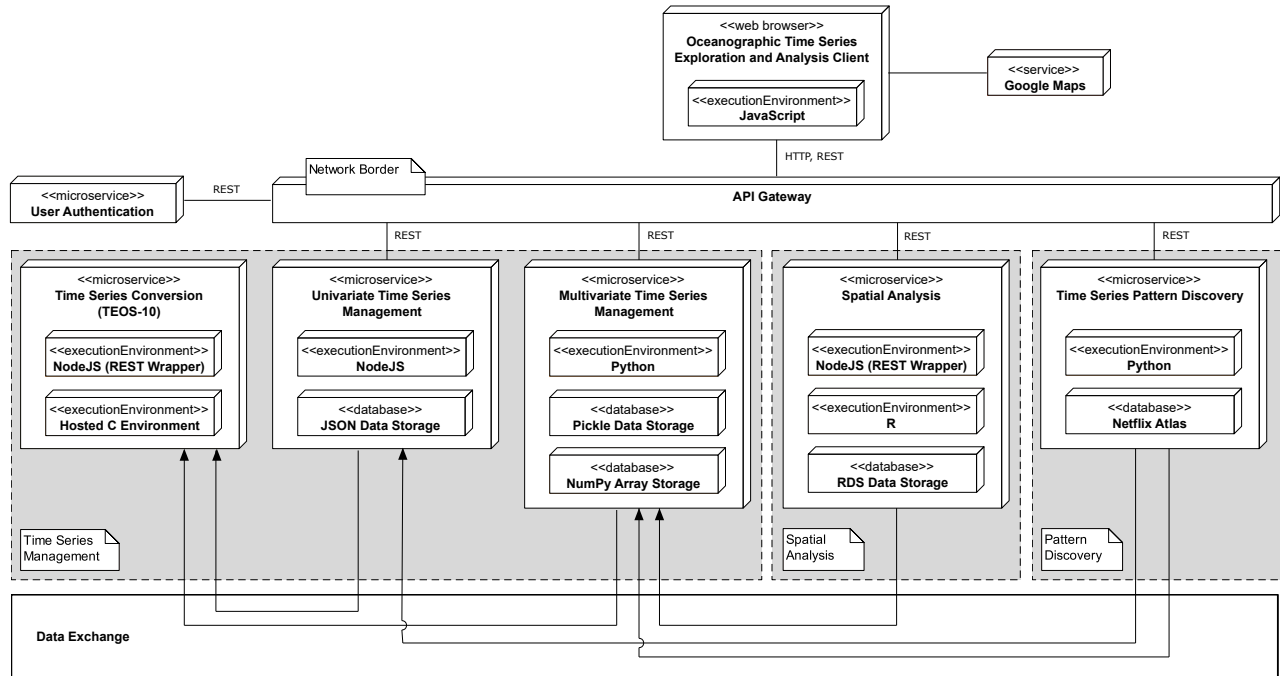


Figure 2. The microservice software architecture of OceanTEA.

an application programming interface (API) gateway. This gateway masks the complexity of communicating with different services and offers an integrated API to the client. The microservices comprising OceanTEA are divided into three so-called *verticals* that group services with related functionality. In the first vertical, we arranged microservices related to the management of time series. In the second vertical, we find the service for the spatial analysis of data, and the third vertical consists of the service for pattern discovery in time. Note, that we made use of the polyglot properties of microservices; for example, we reused an existing C implementation of TEOS-10 for the conversion microservice and implemented the multidimensional array handling of the multivariate time series management in Python, which allows to express the required slicing of arrays with a concise syntax.

We utilize Docker [18] to run each microservice in an isolated container. These containers can directly be deployed to private or public cloud infrastructure. Via Docker Machine [19], the same containers can also be executed on desktop computers running Mac OS X or Microsoft Windows. For these two platforms, we built installer applications to make the installation process user-friendly.

IV. FUTURE WORK

In the future, we plan to extend the spatial analysis view of OceanTEA with interactive 3D renderings of

data fields in relation to structures on the ocean floor (work in progress). For the temporal analysis view, we are working on an implementation that leverages machine learning methods [20] to identify dependencies between different (lagged) time series.

OceanTEA will be used to interactively visualize, explore, and analyze oceanographic time series data for climate-relevant research concerning ocean physics, biology, and chemistry in a changing climate system. For example, OceanTEA is currently employed in studying the impact of ongoing climate change on cold-water coral reef ecosystems in order to assess whether factors such as ocean warming and acidification impact the physical and biogeochemical boundary conditions of these reefs [cf. 21]. We implemented an interactive illustration of modeling results with OceanTEA to accompany a publication which we prepared in this context [22]. In this way, our tool can be used to create dynamic visualizations of figures in papers to add value to publications reporting on data-driven research right from the beginning of the peer review process.

ACKNOWLEDGMENTS

This project was funded by the Cluster of Excellence 80 “The Future Ocean.” The “Future Ocean” is funded within the framework of the Excellence Initiative by the Deutsche Forschungsgemeinschaft (DFG) on behalf of the German federal and state governments.

REFERENCES

- [1] D. Roemmich, G. C. Johnson, S. C. Riser, R. E. Davis, J. Gilson, W. B. Owens, S. L. Garzoli, C. Schmid, and M. Ignaszewski, “The argo program: Observing the global ocean with profiling floats,” *Oceanography*, vol. 22, pp. 34–43, 2009.
- [2] L. Rovelli, K. M. Attard, L. D. Bryant, S. Flögel, H. J. Stahl, M. Roberts, P. Linke, and R. N. Glud, “Benthic O₂ uptake of two cold-water coral communities estimated with the non-invasive eddy-correlation technique,” *Marine Ecology Progress Series*, vol. 525, pp. 97–104, 2015.
- [3] S. Flögel and W. Dullo, “High-resolution water mass measurements around cold-water corals: a comparative test study between repeated Conductivity-Temperature-Depth (CTD) casts and continuous data acquisition of bottom waters from the West Florida Slope, Gulf of Mexico,” *Annalen des Naturhistorischen Museums in Wien. Serie A für Mineralogie und Petrographie, Geologie und Paläontologie, Anthropologie und Prähistorie*, pp. 209–224, 2011.
- [4] S. Newman, *Building Microservices*. O’Reilly, 2015.
- [5] E. Wolff, *Microservices: Flexible Software Architectures*. CreateSpace, 2016.
- [6] H. A. Piwowar, T. J. Vision, and M. C. Whitlock, “Data archiving is a good investment,” *Nature*, vol. 473, no. 7347, pp. 285–285, 2011.
- [7] The Apache Software Foundation, “Apache license, version 2.0.” <https://www.apache.org/licenses/LICENSE-2.0>, 2004.
- [8] R. Schlitzer, “Ocean Data View.” <http://odv.awi.de>, 2016.
- [9] Google Inc., “Google Maps.” <https://maps.google.com>, 2016.
- [10] A. N. Johanson, “CanvasPlot: A JavaScript plotting library based on D3.js for visualizing large data sets.” <https://github.com/a-johanson/canvas-plot>, 2016.
- [11] M. Bostock, “D3.js: Data-Driven Documents.” <https://d3js.org>, 2016.
- [12] T. J. McDougall and P. M. Barker, “Getting started with TEOS-10 and the Gibbs Seawater (GSW) oceanographic toolbox,” tech. rep., SCOR/IAPSO Working Group 127, 2011.
- [13] J. Thönes, “Microservices,” *IEEE Software*, vol. 32, no. 1, pp. 113–116, 2015.
- [14] E. Evans, *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley, 2004.
- [15] W. Hasselbring, “Microservices for scalability,” in *Proceedings of the 7th ACM/SPEC on International Conference on Performance Engineering*, pp. 133–134, 2016.
- [16] R. von Massow, A. van Hoorn, and W. Hasselbring, “Performance simulation of runtime reconfigurable component-based software architectures,” in *Software Architecture (Proceedings ECSA 2011)*, vol. 6903 of *Lecture Notes in Computer Science*, pp. 43–58, Springer, 2011.
- [17] A. van Hoorn, M. Rohr, I. A. Gul, and W. Hasselbring, “An adaptation framework enabling resource-efficient operation of software systems,” in *Proc. of the Warm Up Workshop (WUP 2009) for ACM/IEEE ICSE 2010*, pp. 37–40, 2009.
- [18] Docker Inc., “Docker.” <https://www.docker.com>, 2016.
- [19] Docker Inc., “Docker Machine.” <https://www.docker.com/products/docker-machine>, 2016.
- [20] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- [21] S. Flögel, W. Dullo, O. Pfannkuche, K. Kiriakoulakis, and A. Rüggeberg, “Geochemical and physical constraints for the occurrence of living cold-water corals,” *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 99, pp. 19–26, 2014.
- [22] A. N. Johanson, S. Flögel, W. Dullo, P. Linke, and W. Hasselbring, “Modeling polyp activity of *Paragorgia arborea* using supervised learning.” <https://github.com/a-johanson/paragorgia-arborea-activity>, 2016.